INFORMATION AND CONTROL 4, 169–184 (1961)

Some Remarks on the Capacity of Compound Channels in the Semicontinuous Case*

HARRY KESTEN

The Hebrew University, Jerusalem, Israel

I. INTRODUCTION AND DEFINITIONS

Since we use the standard terminology of coding and information theory as it can be found in Feinstein (1958) or Wolfowitz (1961) we shall be brief in describing the setup. Consider a situation where a sender can transmit n symbols over a (noisy) channel s. The symbols are to be chosen from an input alphabet which is assumed to be the set $\{1, 2, \cdots, a\}$ for all channels under consideration. The channel s may be any one from a given set S and remains the same for all n letters (this is the meaning of the term compound channel, this name being introduced in Wolfowitz, 1961). The choice of the transmission channel cannot be influenced by sender or receiver but in some circumstances (cf. Section III) may be known to one or both of them. The symbols received by the receiver belong to an output alphabet which may depend on s but which (by definition of the term semicontinuous) may be infinite. In order to make life easier we assume the output alphabet to be the set of integers for all $s \in S$. Theorems 1, 4, and the first part of Theorem 3, however, carry over without difficulty to the more general setup described by Feinstein (1958) and Wolfowitz (1961) where the output alphabet belongs to any space with a given Borel field.

If a sequence $u = (i_1, \cdots, i_n)$ of n letters is transmitted, the received sequence of n letters, say, $v(u) = (Y_1(u), \cdots, Y_n(u))$ is a random variable. We assume the channels in S to be stationary, memoryless, and without anticipation, i.e., there exist channel probability functions

$$w(j \mid i \mid s) \geqq 0, \qquad \sum_j w(j \mid i \mid s) = 1 \qquad (i = 1, \cdots, a; s \in S)$$

* Research supported in part by project NONR 266(04) at Columbia University, New York, N. Y.

such that[1]

$$P\{Y_1(u) = j_1, \cdots, Y_n(u) = j_n \mid u \mid s\},$$

the probability of receiving $(j_1, \cdots, j_n)$ when $u = (i_1, \cdots, i_n)$ is transmitted over s, equals $\prod_{k=1}^{n} w(j_k \mid i_k \mid s)$.

The definition of a code depends on the knowledge of the channel by the sender and or receiver. If neither knows the channel over which the message is transmitted, a (n, N, λ) code for the compound channel is defined as a set

$$\{(u_1, A_1), \cdots, (u_N, A_N)\}$$

where each u_i is a sequence $(i_1, \cdots, i_n)(i_j = 1, 2, \cdots, a)$ and the A_i are disjoint sets of sequences $(y_1, \cdots, y_n)$ of n integers, such that

$$P\{v(u_i) \in A_i \mid u_i \mid s\} \geqq 1 - \lambda \qquad (i = 1, \cdots, N; s \in S). \quad (1.1)$$

The u_i and A_i do not depend on s. If the sender only knows the channel of transmission, the u_i's but not the A_i may depend on s. A (n, N, λ) code

$$\{(u_1(s), A_1), \cdots, (u_N(s), A_N)\}$$

must now satisfy

$$P\{v(u_i(s)) \in A_i \mid u_i(s) \mid s\} \geqq 1 - \lambda \qquad (i = 1, \cdots, N; s \in S). \quad (1.2)$$

If the receiver only knows the channel, the A_i but not the u_i may depend on s and A_i in (1.1) is replaced by $A_i(s)$, while finally if both sender and receiver know the channel, u_i and A_i may depend on s. The possible use of such codes has been described by Wolfowitz, 1960.

We put

$N_1(n, \lambda)$ = maximal N for which a (n, N, λ) code exists for S if neither sender nor receiver knows the channel over which the message is transmitted.

$N_i(n, \lambda)$ for $i = 2, 3, 4$ are the maximal N for which a (n, N, λ) code exists respectively if the sender only $(i = 2)$, the receiver only $(i = 3)$, the sender and receiver $(i = 4)$ know the channel.

[1] If we use symbols like $w(a|b|c)$ or $P\{a|b|c\}$ a will always specify an output, b an input, and c the channel and the symbol will be "the probability to receive a (or an output specified by a as in (1.1)) if b is transmitted over the channel specified by c."

Finally we want to define $I(\Pi)$. If any channel with input alphabet $\{1, \cdots, A\}$, output alphabet the integers, and channel probability function

$$w(j \mid i) = \text{probability of receiving } j \text{ if } i \text{ is transmitted,}$$

is given, we define

$$I(\Pi) = \sum_{i=1}^{A} \sum_{j} \Pi(i) w(j \mid i) \log w(j \mid i)(w(j))^{-1}. \tag{1.3}$$

Here $\Pi = (\Pi(1), \cdots, \Pi(A))$ is any probability vector on the input alphabet ($\Pi(i) \geqq 0$, $\sum_{i=1}^{A} \Pi(i) = 1$) while

$$w(j) = \sum_{i=1}^{A} \Pi(i) w(j \mid i).$$

$I(\Pi)$ is usually introduced as the difference of two entropies— the entropy of an output letter minus the expected conditional entropy of an output letter given the input letter. Equation (1.3) can be rewritten as

$$\begin{aligned} I(\Pi) &= -\sum_{i=1}^{A} \Pi(i) \log \Pi(i) \\ &\quad + \sum_{j} w(j) \sum_{i=1}^{A} \Pi(i) w(j \mid i)(w(j))^{-1} \log \Pi(i) w(j \mid i)(w(j))^{-1}. \end{aligned} \tag{1.4}$$

Since

$$\sum_{i=1}^{A} \left| \Pi(i) w(j \mid i)(w(j))^{-1} \log \Pi(i) w(j \mid i)(w(j))^{-1} \right| \leqq \log A$$

(cf. Theorem 1, p. 15 in Feinstein, 1958), the series in (1.3) converge absolutely. In addition one has the well known inequality

$$0 \leqq I(\Pi) \leqq \log A \tag{1.5}$$

(cf. p. 26 in Feinstein, 1958). We shall write $I(\Pi \mid s)$ if we consider the above function for the channel s (in which case A has to be taken as a and $w(j \mid i)$ is replaced by $w(j \mid i \mid s)$.

It was proved by Wolfowitz (1960) that if the output alphabet for all $s \in S$ is the *same, finite* set $\{1, \cdots, b\}$ then, for any $0 < \lambda < 1$,[2]

[2] In this note all logarithms are to the base 2 and 0log0 is taken to be zero.

$$\lim_{n\to\infty} n^{-1} \log N_1(n, \lambda) = \lim_{n\to\infty} n^{-1} \log N_3(n, \lambda) = \max_{\Pi} \inf_{s\epsilon S} I(\Pi \mid s) = C_0 \text{, say,} \tag{1.6}$$

and

$$\lim_{n\to\infty} n^{-1} \log N_2(n, \lambda) = \lim_{n\to\infty} n^{-1} \log N_4(n, \lambda) = \inf_{s\epsilon S} \max_{\Pi} I(\Pi \mid s) = C_1 \text{, say,} \tag{1.7}$$

(Π runs through all probability vectors on $\{1, \cdots, a\}$). We shall start in Section II with an example due to J. H. B. Kemperman to show that (1.6) and (1.7) do not hold in the semicontinuous case. We shall define a number C_2 such that for all $0 < \lambda < 1$

$$\liminf_{n\to\infty} n^{-1} \log N_1(n, \lambda) \geqq C_2 \tag{1.8}$$

while

$$\limsup_{n\to\infty} n^{-1} \log N_1(n, \lambda) \leqq C_2(1 - \lambda)^{-1}. \tag{1.9}$$

In general, $C_2 < C_0$, but in special cases C_2 may equal C_0 (Theorem 2).

Equations (1.8) and (1.9) show that C_2 is the capacity of the compound channel (cf. however, the remarks after Theorem 1). In distinction to (1.6) and (1.7) the knowledge of the channel does influence the maximal code length as is shown in Section III where $N_i(n, \lambda)$ is considered for $i = 2, 3, 4$.

II. THE CODING THEOREM AND ITS CONVERSE

The following example, due to J. H. B. Kemperman, shows that (1.6) and (1.7) do not hold in the semicontinuous case.

For any integer $k \geqq 1$ let

$$E_1^{(k)}, E_2^{(k)}, \cdots, E_{\binom{2k}{k}}^{(k)}$$

be the $\binom{2k}{k}$ different subsets of size k of $\{1, 2, \cdots, 2k\}$. For all $s \epsilon S$ the input alphabet will consist of the two symbols 1,2. For $k \geqq 1$, $r \leqq \binom{2k}{k}$ we define a channel $s(k, r)$ with output alphabet $\{1, 2, \cdots, 2k\}$

by its channel probability function

$$w(j \mid 1 \mid s(k, r)) = \begin{cases} k^{-1} \text{ if } j \in E_r^{(k)} \\ 0 \text{ otherwise} \end{cases}$$

and

$$w(j \mid 2 \mid s(k, r)) = \begin{cases} 0 \text{ if } j \in E_r^{(k)} \text{ or } j > 2k \\ k^{-1} \text{ otherwise.} \end{cases}$$

Thus if a 1 (resp. 2) is transmitted the output letter is uniformly distributed over $E_r^{(k)}$ ($\{1, 2, \cdots, 2k\} - E_r^{(k)}$). S is the collection of all channels obtained in this way $\left(k = 1, 2, \cdots, r = 1, 2, \cdots, \binom{2k}{k}\right)$. If n letters are transmitted over $s(k, r)$ the probability of receiving a given output letter more than once is arbitrary small for sufficiently large k. If the receiver does not know which channel has been used and no output letter appeared more than once, it is intuitively clear that he cannot deduce anything about the transmitted sequence. From Theorems 1 and 3 one can indeed deduce that for this system S

$$N_1(n, \lambda) \leqq (1 - \lambda)^{-1}$$

and also

$$N_2(n, \lambda) \leqq (1 - \lambda)^{-1}.$$

On the other hand one easily verifies that $I(\tilde{\Pi} \mid s(k, r)) = 1$ for $\tilde{\Pi} = (\frac{1}{2}, \frac{1}{2})$ and thus $C_0 = C_1 = 1$. Hence neither (1.6) nor (1.7) holds for the semicontinuous case.

In order to define C_2 for a general class S of channels we consider probability vectors $\Pi^{(k)}$ on the sequences $u = (i_1, \cdots, i_k)$ of k input letters ($i_j = 1, \cdots, a$; $\Pi^{(k)}(u) \geqq 0$; $\sum_{i_1, \cdots, i_k} \Pi^{(k)}(u) = 1$). We also divide the class $\Omega^{(k)}$ of sequences $(y_1, \cdots, y_k)$ (y integer) into a finite number of disjoint sets, $B_1, \cdots, B_r$ say. Let $D^{(k)}$ stand for the dissection $B_1, \cdots, B_r$ of $\Omega^{(k)}$ and introduce

$$\tilde{w}(p \mid u \mid s) = p\{v(u) \in B_p \mid u \mid s\} = \sum_{(j_1, \cdots, j_k) \in B_p} \prod_{t=1}^{k} w(j_t \mid i_t \mid s), \tag{2.1}$$

$$\tilde{w}(p \mid s) = \sum_{u} \Pi^{(k)}(u) \tilde{w}(p \mid u \mid s), \tag{2.2}$$

and

$$I(\Pi^{(k)}|\, D^{(k)}|\, s) = \sum_{u} \sum_{p=1}^{r} \Pi^{(k)}(u)\tilde{w}(p \mid u \mid s) \log \tilde{w}(p \mid u \mid s)(\tilde{w}(p \mid s))^{-1}. \tag{2.3}$$

The summation over u in (2.2) and (2.3) is over all sequences $u = (i_1, \cdots, i_k)$ with $1 \leqq i_j \leqq a$. Note that $I(\Pi^{(k)}|\, D^{(k)}|s)$ is exactly what one would obtain in (1.3) for a channel with input alphabet the a^k sequences $(i_1, \cdots, i_k)$, output alphabet the set $\{1, \cdots, r\}$, and channel probability function $\tilde{w}(p \mid u \mid s)$.

LEMMA 1.[3]

$$C_2 = \lim_{k\to\infty} k^{-1} \sup_{\Pi^{(k)}} \sup_{D^{(k)}} \inf_{s} I(\Pi^{(k)}|\, D^{(k)}|\, s) \tag{2.4}$$

exists and moreover for every k

$$k^{-1} \sup_{\Pi^{(k)}} \sup_{D^{(k)}} \inf_{s} I(\Pi^{(k)}|\, D^{(k)}|\, s) \leqq C_2 .$$

PROOF. Let us write for brevity

$$I_k = \sup_{\Pi^{(k)}} \sup_{D^{(k)}} \inf_{s} I(\Pi^{(k)}|\, D^{(k)}|\, s).$$

Let $(\Pi^{(k_1)}, \Pi^{(k_2)})$ denote the product measure on the sequences $(i_1, \cdots, i_{k_1}, i_{k_1+1}, \cdots, i_{k_1+k_2})$ of the measure $\Pi^{(k_1)}$ for $(i_1, \cdots, i_{k_1})$ and $\Pi^{(k_2)}$ for $(i_{k_1+1}, \cdots, i_{k_1+k_2})$. Similarly, if $D^{(k_1)}(D^{(k_2)})$ dissects $\Omega^{(k_1)}(\Omega^{(k_2)})$ into $B_1, \cdots, B_r(C_1, \cdots, C_s)$ then $(D^{(k_1)}, D^{(k_2)})$ is the dissection of $\Omega^{(k_1+k_2)}$ into the r, s sets (B_i, C_j) consisting of all sequences $(y_1, \cdots, y_{k_1+k_2})$ with $(y_1, \cdots, y_{k_1}) \in B_i$ and $(y_{k_1+1}, \cdots, y_{k_1+k_2}) \in C_j$. It is well known that

$$I((\Pi^{(k_1)}, \Pi^{(k_2)})|(D^{(k_1)}, D^{(k_2)})|\, s) = I(\Pi^{(k_1)}|\, D^{(k_1)}|\, s) + I(\Pi^{(k_2)}|\, D^{(k_2)}|s)$$

since $(i_1, \cdots, i_{k_1})$ and $(i_{k_1+1}, \cdots, i_{k_1+k_2})$ are independent under the distribution $(\Pi^{(k_1)}, \Pi^{(k_2)})$ (cf. Lemma 3, p.15, lemma 6, p. 16, and theorem on p. 29 of Feinstein, 1958). Hence

$$I_{k_1+k_2} \geqq I_{k_1} + I_{k_2}$$

and consequently (cf. Polya and Szego, 1954)

$$\lim_{k\to\infty} k^{-1} I_k = C_2 \text{ exists},$$

[3] In $\inf_s$ s will always range over S.

and

$$k^{-1}I_k \leqq C_2 \quad \text{for all } k.$$

C_2 is finite by (1.5).

THEOREM 1. *For any* $0 < \lambda < 1$

$$\liminf_{n\to\infty} n^{-1} \log N_1(n, \lambda) \geqq C_2 \tag{2.5}$$

and for any $n \geqq 1$

$$\log N_1(n, \lambda) \leqq (nC_2 + 1)(1 - \lambda)^{-1}. \tag{2.6}$$

PROOF: Given any $\varepsilon > 0$ we can find a k_0, $\tilde{\Pi}^{(k_0)}$, $\tilde{D}^{(k_0)}$ such that

$$\inf_s I(\tilde{\Pi}^{(k_0)} | \tilde{D}^{(k_0)} | s) \geqq k_0(C_2 - \tfrac{1}{2}\varepsilon).$$

Let $\tilde{D}^{(k_0)} = \{B_1, \cdots, B_r\}$. Consider now the channels of S as channels with input alphabet all sequences $u = (i_1, \cdots, i_{k_0})$ with $i_j = 1, \cdots, a$, as output alphabet the set $\{1, \cdots, r\}$ and with channel probability function

$$\tilde{w}(p \mid u \mid s) = P\{v(u) \in B_p \mid u \mid s\}.$$

Since all these channels have the same finite output alphabet (1.5) is applicable with C_0 replaced by

$$\sup_{\Pi^{(k_0)}} \inf_s I(\Pi^{(k_0)} | \tilde{D}^{(k_0)} | s) \geqq k_0(C_2 - \tfrac{1}{2}\varepsilon).$$

Consequently, there exists for any $0 < \lambda < 1$ and sufficiently large m a $(m, [2^{mk_0(C_2-\varepsilon)}], \lambda)$ code for these channels. Any such code is obviously a $(mk_0, [2^{mk_0(C_2-\varepsilon)}], \lambda)$ code for S. For such a code, the decoding of a sequence $(y_1, \cdots, y_{k_0}, y_{k_0+1}, \cdots, y_{2k_0}, \cdots, y_{mk_0})$ depends only on the sets B to which $(y_1, \cdots, y_{k_0})$, $(y_{k_0+1}, \cdots, y_{2k_0})$, $\cdots$, $(y_{(m-1)k_0+1}, \cdots, y_{mk_0})$ belong. Using the fact that $N_1(n, \lambda)$ is increasing in n, (2.5) follows. As for (2.6), assume that

$$\{(u_1, A_1), \cdots, (u_N, A_N)\}$$

is a (n, N, λ) code. Without loss of generality we may assume that $\bigcup_{i=1}^{N} A_i = \Omega^{(n)}$. Hence $A_1, \cdots, A_N$ is a finite dissection say $D^{(n)}$ of $\Omega^{(n)}$. In addition, for every $s \in S$

$$P\{v(u_i) \in A_i \mid u_i \mid s\} \geqq 1 - \lambda.$$

Define $\Pi^{(n)}$ by

$$\Pi^{(n)}(u_i) = N^{-1} \qquad i = 1, \cdots, N$$
$$\Pi^{(n)}(u) = 0 \qquad \text{for any sequence } (i_1, \cdots, i_n) \text{ not equal to } u_1 \text{ or } u_2 \cdots \text{ or } u_N .$$

It was proved by Feinstein (1958, pp. 35 and 44) and by Wolfowitz (1961, Section 7.4), that in this case

$$(1 - \lambda) \log N \leqq I(\Pi^{(n)} \mid D^{(n)} \mid s) + 1.$$

This holds for all s and thus

$$(1 - \lambda) \log N_1(n, \lambda) \leqq \inf_s I\,(\Pi^{(n)} \mid D^{(n)} \mid s) + 1 \leqq I_n + 1 \leqq nC_2 + 1$$

which completes the proof.

REMARK 1. Equations (2.5) and (2.6) prove in the terminology of Blackwell *et al.* (1959) that C_2 is the supremum of all attainable rates of transmission. Wolfowitz (1960) called (2.6) the weak converse of the coding theorem while one would have to prove the strong converse, i.e., $\lim \sup_{n\to\infty} n^{-1} \log N_1(n, \lambda) \leqq C_2$ in order to call C_2 the capacity of S. The author has been unable to prove the strong converse. The difficulty lies in the fact that, in general, $C_2 < C_0$ (e.g., for the example above $C_0 = 1$ while one can show $C_2 = 0$). This makes the approach of Wolfowitz (1960) to prove the strong converse by looking at individual channels inapplicable.

REMARK 2. One would of course like to have estimates on the speed of convergence of $n^{-1}I_n$ to C_2. Unfortunately no estimates independent of the system S exist. For example, if the input alphabet consists of $\{1, 2\}$ for all channels in S one can always add all channels $s(k, r)$ with $k \leqq M$ from the example at the beginning of this section without reducing C_2. For any fixed n, however, I_n will tend to zero if M tends to infinity. It seems to the author that this nonuniformity in S is a serious drawback. It makes it a.o. very hard if not impossible to find C_2 for many systems.

We shall now describe a situation where $C_2 = C_0$. Before phrasing the next lemma let us remark that the output alphabet of s is actually the set of all j for which

$$\sum_{i=1}^{a} w(j \mid i \mid s) > 0.$$

For convenience we may sometimes add some j's with $\sum_{i=1}^{a} w(j \mid i \mid s) = 0$.

LEMMA 2. *If the output alphabets of the channels in S are uniformly bounded in size, then*

$$\liminf_{n\to\infty} n^{-1} \log N_1(n, \lambda) \geqq \sup_{\Pi} \inf_{s} I(\Pi \mid s) = C_0 .$$

(This is a generalization of (1.6) only if the union of all output alphabets is infinite).

PROOF. Let the upper bound of the sizes of the output alphabets be b. By adding, if necessary, some letters with $\sum_{j=i}^{a} w(j \mid i \mid s) = 0$ we may assume that each channel $s \in S$ has an output alphabet of exactly b letters which we shall denote by

$$j_1(s), \cdots, j_b(s).$$

If σ is any permutation of $\{1, \cdots, b\}$ we define a channel $\sigma(s)$ with transition probability function

$$w(k \mid i \mid \sigma(s)) = w(j_{\sigma^{-1}(k)}(s) \mid i \mid s) \qquad k = 1, \cdots, b$$

and

$$w(k \mid i \mid \sigma(s)) = 0 \qquad \text{for} \quad k > b.$$

(i.e., $\sigma(s)$ is the channel obtained from s by mapping the output letters onto the set $\{1, \cdots, b\}$ and by permuting this set). Obviously,

$$I(\Pi \mid \sigma(s)) = I(\Pi \mid s)$$

and hence, if

$$S' = \{\sigma(s) : s \in S, \sigma \text{ any permutation of } \{1, \cdots, b\}\}$$

then

$$\sup_{\Pi} \inf_{\sigma(s) \in S'} I(\Pi \mid \sigma(s)) = \sup_{\Pi} \inf_{s} I(\Pi \mid s) = C_0 .$$

All the channels in S' have the same output alphabet $\{1, \cdots, b\}$ so that (1.6) is again applicable. Thus for any $\varepsilon > 0$, $0 < \lambda < 1$, and sufficiently large n, there exists for S' a

$$(n, [2^{n(C_0-\varepsilon)}], \lambda) \tag{2.7}$$

code. Let (with $N = [2^{n(C_0-\varepsilon)}]$)

$$\{(u_1, A_1'), \cdots, (u_N, A_N')\} \tag{2.8}$$

be such a code where the A_i' are disjoint sets of sequences $(y_1, \cdots, y_N)$ $(y_i = 1, \cdots, b)$ which are symmetric in $1, \cdots, b$, i.e., if $(y_1, \cdots, u_n) \in A_i'$ then also $(\sigma(y_1), \sigma(y_2), \cdots, \sigma(y_n)) \in A_i'$ for any permutation σ of $\{1, \cdots, b\}$. That we can choose A_i' symmetric can be seen from the construction by Wolfowitz (1960, Section 4) of these codes. As the output letters $1, \cdots, b$ all play the same role for S' it is possible to do every step of the construction in a symmetric way. If now $(y_1, \cdots, y_n)$ is any sequence of n integers, containing at most b different elements, say

$$j_1 < j_2 < \cdots < j_c \qquad (c \leqq b)$$

then we put

$$\tau(y_1, \cdots, y_n) = (z_1, \cdots, z_n) \quad \text{where} \quad z_i = k \quad \text{if} \quad y_i = j_k .$$

As a code for the system S we now take

$$\{(u_1, A_1), \cdots, (u_N, A_N)\} \tag{2.9}$$

where

$$A_i = \{(y_1, \cdots, y_n) : \tau(y_1, \cdots, y_n) \in A_i'\}.$$

Clearly the A_i are disjoint and if $j_1(s) < j_2(s) < \cdots < j_b(s)$ is the output alphabet of the channel which happens to be used for the transmission, then the output sequences will be of the form $(y_1, \cdots, y_n)$ with $y_i = j_1(s)$ or $j_2(s)$ $\cdots$ or $j_b(s)$. Moreover, if σ_0 is the identity permutation of $\{1, \cdots, b\}$ and σ_1 any permutation mapping k_1 on 1, k_2 on $2, \cdots, k_c$ on c $(k_1 \leqq k_2 \cdots \leqq k_c\,;\, c \leqq b)$ then one has

$$\begin{aligned}
&P\left\{\begin{array}{l|c|c|}\text{the received sequence } (y_1, \cdots, y_n) \in A_i & u_i & \\ \text{and contains only } j_{k_1}(s) < j_{k_2}(s) < \cdots & \text{transmitted} & s \\ \qquad\qquad < j_{k_c}(s) & & \end{array}\right\} \\
&= P\left\{\begin{array}{l|c|c|}\tau(y_1, \cdots, y_n) \in A_i' \text{ and contains} & u_i & s \\ \text{only } 1, 2, \cdots, c & & \end{array}\right\} \\
&= P\left\{\begin{array}{l|c|c|}\text{received sequence} \in A_i' \text{ and} & u_i & \sigma_1(s) \\ \text{contains only } 1, 2, \cdots, c & & \end{array}\right\} \\
&= P\left\{\begin{array}{l|c|c|}\text{received sequence} \in A_i' \text{ and} & u_i & \sigma_0(s) \\ \text{contains only } k_1, \cdots, k_c & & \end{array}\right\}.
\end{aligned} \tag{2.10}$$

In the last equality the symmetry of A_i' in $1, \cdots, b$ was used. From

(2.10) we have immediately

$$P\{v(u_i) \in A_i \mid u_i \mid s\} = P\{v(u_i) \in A_i' \mid u_i \mid \sigma_0(s)\} \geqq 1 - \lambda \quad (2.11)$$

which proves that (2.9) is a (n, N, λ) code for S. This together with (2.7) proves the lemma.

Let us now return to the more general situation of infinite output alphabets. Rank the j's according to decreasing size of w, i.e., let $j_1(i, s)$, $j_2(i, s)$, $\cdots$ be a permutation of the integers such that

$$w(j_1(i, s) \mid i \mid s) \geqq w(j_2(i, s) \mid i \mid s) \geqq \cdots$$

and put

$$r_m(i \mid s) = \sum_{k=1}^{m} w(j_k(i, s) \mid i \mid s).$$

r_m is the sum over the k largest probabilities of output letters.

THEOREM 2. *If*

$$\liminf_{m \to \infty} \inf_{s} r_m(i \mid s) = 1 \quad (2.12)$$

for $i = 1, \cdots,$ *a then for any* $0 < \lambda < 1$

$$\lim_{n \to \infty} n^{-1} \log N_1(n, \lambda) = \sup_{\Pi} \inf_{s} I(\Pi \mid s) = C_0 .$$

PROOF. We shall start to prove

$$\liminf_{n \to \infty} n^{-1} \log N_1(n, \lambda) \geqq C_0. \quad (2.13)$$

For any $\eta > 0$ we derive a channel $s(\eta)$ from the channel s. For this purpose, let $B(\eta, s)$ be a set of letters such that

$$P\{j \in B(\eta, s) \mid i \mid s\} \geqq 1 - \eta \qquad i = 1, \cdots, a. \quad (2.14)$$

The channel $s(\eta)$ is obtained by "lumping" all output letters which are not in $B(\eta, s)$. Thus the output alphabet of $s(\eta)$ consists of $B(\eta, s)$ plus one symbol $\tilde{B}$ while

$$w(j \mid i \mid s(\eta)) = w(j \mid i \mid s) \qquad \text{for} \quad j \in B(\eta, s)$$

and

$$w(\tilde{B} \mid i \mid s(\eta)) = 1 - \sum_{j \in B(\eta, s)} w(j \mid i \mid s).$$

By (1.3) and (1.4)

$$I(\Pi \mid s(\eta)) = -\sum_{i=1}^{a} \Pi(i) \log \Pi(i) \sum_{j \in B(\eta,s)} w(j \mid i \mid s)$$

$$+ \sum_{j \in B(\eta,s)} w(j \mid s) \sum_{i=1}^{a} \Pi(i) w(j \mid i \mid s)(w(j \mid s))^{-1}$$

$$\cdot \log \Pi(i) w(j \mid i \mid s)(w)j \mid s))^{-1} \quad (2.15)$$

$$+ \sum_{i=1}^{a} \Pi(i) w(\bar{B} \mid i \mid s(\eta)) \log w(\bar{B} \mid i \mid s(\eta))$$

$$\cdot (\sum_{k=1}^{a} \Pi(k) w(\bar{B} \mid k \mid s(\eta))^{-1}.$$

As $\eta \to 0$ this approaches $I(\Pi \mid s)$ uniformly in Π and s, for

$$\sum_{i=1}^{a} | \Pi(i) w(j \mid i \mid s)(w(j \mid s))^{-1} \log \Pi(i) w(j \mid i \mid s)(w(j \mid s))^{-1} |$$

$$\leqq \log a.$$

Consequently, given any $\varepsilon > 0$, we can fix $\eta = \eta(\varepsilon)$ such that

$$\sup_{\Pi} \inf_{s(\eta)} I(\Pi \mid s(\eta)) \geqq C_0 - \tfrac{1}{3}\varepsilon \quad (2.16)$$

where $s(\eta)$ runs through all channels obtainable from some $s \in S$ by any choice of $B(\eta, s)$ satisfying (2.14). By (2.12) there exists an $m(\eta)$ independent of s and sets $B(\frac{1}{2}\eta, s)$ with

$$| B(\tfrac{1}{2}\eta, s)| \leqq m(\eta).$$

($| B |$ denotes the number of elements in B). The sender may now start to transmit first M times the symbol 1, then M times the symbols 2 etc., up to M times the symbols a. The (random) set $C_i(\eta)$ will consist of the $m(\eta)$ output letters which occur most often among those received while i was transmitted M times (if several choices for $C_i(\eta)$ are possible due to equality of some frequencies it does not matter which choice is made). For sufficiently large M, the set

$$\bigcup_{i=1}^{a} C_i(\eta)$$

has a probability of at least $1 - \frac{1}{2}\lambda$ to be a $B(\eta, s)$ set for the channel

which is used for the transmission. Moreover,

$$|\bigcup_{i=1}^{a} C_i(\eta)| \leqq am(\eta).$$

By Lemma 2 and (2.16) there exists for sufficiently large n a

$$(n, [2^{n(C_0-(2\varepsilon/3))}], \lambda)$$

code for the compound system of all obtainable channels $s(\eta)$ which have

$$|B(\eta, s)| \leqq am(\eta).$$

Since the receiver knows such a $B(\eta, s)$ set (i.e., $\bigcup_{i=1}^{a} C_i(\eta)$) with a probability not less than $1 - \frac{1}{2}\lambda$ at the cost of aM letters, it follows that there exists for S a

$$(n, [2^{(n-aM)(C_0-(2\varepsilon/3))}], \lambda)$$

code. This proves (2.13).

In order to prove the strong converse

$$\limsup_{n\to\infty} n^{-1} \log N_1(n, \lambda) \leqq C_0 \tag{2.17}$$

we follow Kemperman (1961) and Wolfowitz (1960). Let $\{(u_1, A_1), \cdots, (u_N, A_N)\}$ be a (n, N, λ) code. Let the input letter i occur $f_{i,j}$ times in u_j. $f_{i,j}$ can take the $(n + 1)$ values $0, 1, \cdots, n$. We can thus divide the u_i into at most $(n + 1)^a$ disjoint subsets such that u_{j_1} and u_{j_2} are in the same subset if and only if

$$f_{i,j_1} = f_{i,j_2} \qquad i = 1, \cdots, a.$$

Let the largest such subset have $N' \geqq (n + 1)^{-a}N$ members. Then there is a (n, N', λ) code $\{(u_{j_1}, A_{j_1}), \cdots, (u_{j_{N'}}, A_{j_{N'}})\}$ such that the frequencies $p_i = n^{-1}f_{i,j_s}$ are the same for $s = 1, \cdots, N'$. It follows from Kemperman (1961) that there exists a constant K independent of s and p such that any such code must satisfy

$$(n + 1)^{-a}N \leqq N' \leqq 2^{nI(p|s)+K\sqrt{n}}. \tag{2.18}$$

Taking the infimum over s, the supremum over p, and letting $n \to \infty$ completes the proof. A special case of Theorem 2 is the case described in Lemma 2 where all channels have output alphabets bounded in size.

III. TRANSMISSION WITH KNOWLEDGE OF THE CHANNEL BY SENDER OR RECEIVER

In Section I we already defined codes for the situation where sender or receiver know the channel of transmission. Such situations were discussed for the first time by Wolfowitz (1960). Using (1.7) instead of (1.6) one can follow the proofs of Theorems 1 and 2 to obtain

THEOREM 3.

$$C_3 = \lim_{k\to\infty} k^{-1} \sup_{D^{(k)}} \inf_{s} \sup_{\Pi^{(k)}} I(\Pi^{(k)} | D^{(k)} | s)$$

exists, and if the sender only knows the channel of transmission, then for any $0 < \lambda < 1$

$$\liminf_{n\to\infty} n^{-1} \log N_2(n, \lambda) \geqq C_3 \tag{3.1}$$

and for any $n \geqq 1$

$$\log N_2(n, \lambda) \leqq (nC_3 + 1)(1 - \lambda)^{-1}. \tag{3.2}$$

If (2.12) *is satisfied, then*

$$\lim_{n\to\infty} n^{-1} \log N_2(n, \lambda) = \inf_{s} \sup_{\Pi} I(\Pi | s) = C_1 . \tag{3.3}$$

The proof of (3.3) requires a bit more care than the proof of Lemma 2 and Theorem 2. For example, where we want to prove the analogue of Lemma 2, the sender may know the channel which is used but he does not know the appropriate permutation σ. Looking at the construction of the code in Wolfowitz (1960, Section 8) we see that the crucial quantity really is $\Pi(s)$ (in the notation of Wolfowitz). However, $I(\Pi \mid \sigma_1(s)) = I(\Pi \mid \sigma_2(s)$ for every Π and pair of permutations σ_1 and σ_2, so that $\Pi(s)$ can be chosen without knowledge of σ. Similarly, for the analogue of Theorem 2, the sender knows the channel but not the set $\bigcup_{i=1}^{a} C_i(\eta)$. We have seen in (2.15), however, that $I(\Pi \mid s) - I(\Pi \mid s(\eta))$ tends to zero as $\eta \to 0$ uniformly in Π, s and $s(\eta)$ and thus $I(\Pi \mid s(\eta))$ hardly depends on the $B(\eta, s)$ set. Consequently the $\Pi(s(\eta))$ required for Wolfowitz' proof (1960) can be chosen without exact knowledge of $B(\eta, s)$ (this requires Wolfowitz' estimates of order $n^{-1/2}$ to be changed into $o(n)$).

The example in Section 2 shows that knowledge of the channel by the receiver can in the semicontinuous case increase the maximal code length. This will be made more precise in the next theorem which was conjectured by J. H. B. Kemperman.

THEOREM 4. *If the receiver only knows the channel, then for all* $0 < \lambda < 1$

$$\lim_{n\to\infty} n^{-1} \log N_3(n, \lambda) = \sup_{\Pi} \inf_{s} I(\Pi \mid s) = C_0 .$$

PROOF. Let Π_0 be fixed such that

$$\inf_{s} I(\Pi_0 \mid s) \geqq C_0 - \tfrac{1}{2}\varepsilon.$$

For each s and m we can split the output letters into m^a disjoint classes

$$E(s, r_1, \cdots, r_a) \qquad 1 \leqq r_i \leqq m,$$

where $E(s, r_1, \cdots, r_a)$ contains those j for which

$$\frac{r_i - 1}{m} < \frac{\Pi_0(i) w(j \mid i \mid s)}{w(j \mid s)} \leq \frac{r_i}{m} \qquad (i = 1, \cdots, m). \quad (3.4)$$

The receiver may of course treat all j's in one such set as indistinguishable. There results a channel, $s(m)$ say, with output alphabet symbols $E(r_1, \cdots, r_a)$ and with channel probability function

$$w(E(r_1, \cdots, r_a) \mid i \mid s(m)) = \sum_{j \in E(s, r_1, \cdots, r_a)} w(j \mid i \mid s).$$

It follows from (3.4) and (1.4) that

$$\lim_{m\to\infty} I(\Pi_0 \mid s(m)) = I(\Pi_0 \mid s) \geqq C_0 - \tfrac{1}{2}\varepsilon$$

uniformly in s. Hence, for sufficiently large m,

$$\inf_{s} I(\Pi_0 \mid s(m)) \geqq C_0 - \varepsilon.$$

Since all channels $s(m)$ (m fixed, $s \in S$) have the same finite output alphabet $\{E(r_1, \cdots, r_a)\}$ and since the receiver, knowing the channel, can decide to which E a received letter belongs, (1.6) is applicable. Thus

$$\liminf_{n\to\infty} n^{-1} \log N_3(n, \lambda) \geqq C_0$$

follows. The converse part follows from the strong converse for one channel as in Theorem 2.

For completeness we remark that

$$\lim_{n\to\infty} n^{-1} \log N_4(n, \lambda) = C_1$$

is entirely trivial even in the semicontinuous case, since if both sides

know the channel, the problem reduces essentially to the problem with one channel only.

ACKNOWLEDGMENTS

The author wishes to thank Professors J. H. B. Kemperman and J. Wolfowitz for many stimulating discussions and suggestions.

REFERENCES

BLACKWELL, DAVID, BREIMAN, LEO AND THOMASIAN, A. J. (1959). The capacity of a class of channels. *Ann. Math. Statist.* **30,** 1229–1241.

FEINSTEIN, AMIEL (1958). "Foundations of Information Theory." McGraw Hill, New York.

KEMPERMAN, J. H. B. (1961). To be published. The result can also be found in Section 8.4 of Wolfowitz (1961). Cf. also. Abstract, *Notices Am. Math. Soc.* **7,** 924 (1960).

POLYA, G. AND SZEGO, G. (1954). "Aufgaben und Lehrsatze aus der Analysis," I. Abschn. Aufgabe 98. Springer, Berlin.

WOLFOWITZ, J. (1960). Simultaneous channels. *Arch. Rational Mech. Analysis 4* 371–386.

WOLFOWITZ, J. (1961). Monograph on coding theory, to appear. Springer, Berlin.